Problems for the Numerical Analysis of the Future

Department of Commerce
National Bureau of Standards
Applied Mathematics Series • 15

Applied Mathematics Program of the National Bureau of Standards

National Applied Mathematics Laboratories

The National Applied Mathematics Laboratories, Dr. J. H. Curtiss, chief, constitute a division of the National Bureau of Standards. The functions of the laboratories, in general terms, are to perform research and provide services in various quantitative fields of mathematics. Specifically, the laboratories undertake to conduct basic research in the fields of mathematics extensively used by the physical and engineering sciences, with special emphasis on the development and utilization of high-speed numerical analysis in these fields. The development and construction of pertinent tools, such as automatic high-speed computing machines and mathematical tables, is part of the program, and a general-purpose computing service specializing in digital methods of calculation is provided. Such services as consultation on special problems in these fields, theoretical and in-service training, and the publication of the results of research are included.

Applied Mathematics Series

This series is intended to serve as a vehicle for the publication of mathematical tables, manuals, and studies by the mathematics laboratories of the National Bureau of Standards. The mathematical tables in this series are a continuation of those in the Mathematical Tables series, designated below by MT numbers. The AMS numbers are available from the Superintendent of Documents, Government Printing Office, Washington 25, D. C. Payment should be made to the Superintendent of Documents, by check, money order, or Superintendent of Documents coupons. The prices are for delivery in the United States and its possessions and in countries extending the franking privilege, that is, Canada and most of the Latin-American countries. To other countries, one-third of the list price of the publications should be included in the remittance to cover cost of mailing.

AMS1. Tables of the Bessel Functions $Y_0(x)$, $Y_1(x)$, $K_0(x)$, $K_1(x)$, $0 \leq x \leq 1$. (1948) 60p. 35¢.

AMS2. Table of Coefficients for Obtaining the First Derivative Without Differences. (1948) 20p. 15¢.

AMS3. Tables of the Confluent Hypergeometric Function $F(n/2, 1/2; x)$ and Related Functions. (1949) 73p. 35¢.

AMS4. Tables of Scattering Functions for Spherical Particles. (1948) 119p. 45¢.

AMS5. Table of Sines and Cosines to Fifteen Decimal Places at Hundredths of a Degree. (1949) 95p. 45¢.

AMS6. Tables of the Binomial Probability Distribution. (1950) 387p. $2.50 (blue buckram).

AMS7. Tables to Facilitate Sequential t-Tests. 82p. 45¢.

AMS8. Tables of Powers of Complex Numbers. (1950) 44p. 25¢.

AMS10. Tables for Conversion of X-ray Diffraction Angles to Interplanar Spacing. (1950) 159p. $1.75.

AMS11. Table of Arctangents of Rational Numbers. 105p. $1.50.

AMS12. The Monte Carlo Method. 42p. 30¢.

AMS15. Problems for the Numerical Analysis of the Future. (1951) 21p. 20¢.

Mathematical Tables

A total of 37 mathematical tables have appeared in the MT series of the National Bureau of Standards: Those listed below are available from the Superintendent of Documents, Government Printing Office, Washington 25, D. C. Payment should be made as indicated above for the Applied Mathematics Series.

MT3. Tables of Circular and Hyperbolic Sines and Cosines for Radian Arguments: sin x, cos x, sinh x, cosh x, for $x=0(.0001)2$, 9D. (2d ed. 1949) xvii+405p. $2.50.

MT4. Tables of Sines and Cosines for Radian Arguments: 0(.001)25, 8D. (1940) xix+275p. $2.

MT5. Tables of Sine, Cosine, and Exponential Integrals, Volume I: 0(.0001)2, 9D. (1940) xxvi+444p. $2.75.

MT6. Tables of Sine, Cosine, and Exponential Integrals, Volume II: 0(.001)10; 9S, 10S, or 11S. (1940) xxxvii+225p. $2.

MT8. Tables of Probability Functions, Volume I: 0(.0001)1(.001)5.6, 15D. (1941) xxviii+302p. $2.

MT9. Table of Natural Logarithms, Volume II: 50,000(1)100,000, 16D. (1941) xviii+501p. $3.

MT11. Tables of the Moments of Inertia and Section Moduli of Ordinary Angles, Channels, and Bulb Angles With Certain Plate Combinations. (1941) xiii+197p. $2.

MT13. Table of Sine and Cosine Integrals for Arguments from 10 to 100. (1942) xxxii+185p. $2.

[List continued on cover page 3]

UNITED STATES DEPARTMENT OF COMMERCE • Charles Sawyer, *Secretary*

NATIONAL BUREAU OF STANDARDS • E. U. Condon, *Director*

Problems for the Numerical Analysis of the Future

National Bureau of Standards
Applied Mathematics Series • 15

Issued June 29, 1951

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1951
FOR SALE BY THE SUPERINTENDENT OF DOCUMENTS, U. S. GOVERNMENT PRINTING OFFICE, WASHINGTON 25, D. C.
PRICE 20 CENTS

Foreword

It has been recognized by many persons engaged in the development of large-scale automatic digital computing equipment that the demands of such machines upon men will be very much greater than the demands of men upon the machines. In other words, skill in the analysis, formulation, and programing of problems will become the controlling factor in the proper use of the computing machines of the future.

To attack this problem at its roots, the Institute for Numerical Analysis of the National Bureau of Standards was established early in 1948 as an integral part of the Bureau's extensive applied mathematics and machine development program. The program of the research staff of the Institute consists in the examination of new, and the reexamination of old problems in mathematics with a view toward devising the numerical techniques most suitable for their solution on high-speed automatic computers. A computation laboratory at the Institute, containing the most modern equipment available, is used part time to perform experimental calculations for the research staff. A new computational science, founded partly on the older hand-machine techniques and partly on theories radically new in numerical analysis, is now taking shape at the Institute and elsewhere. It will form a reservoir of knowledge that can be drawn upon for innumerable applications by the users of machines.

Shortly after the Institute was established, a series of symposia on the development of high-speed automatic computing machinery and related numerical methods was held on the campus of the University of California, Los Angeles, where the Institute is located. The symposia served as dedicatory exercises for the Institute. The interest of the scientific public in these topics is indicated by the fact that approximately 500 persons registered for the meetings. Most of the papers given were in the nature of progress reports on the various machine development projects in the United States and in Great Britain. On the last day a series of mathematical papers dealing with the future of numerical analysis was given.

Four of the papers given at the mathematical sessions are presented here. The first one gives the reader a glimpse into the workshop of one of the foremost numerical analysts of our day. It concludes with some penetrating remarks concerning the psychological difficulties involved in trying to harness a robot which can perform in a flash something which used to take a hundred hand computers a year to do. The other three papers deal with problems which have proved inaccessible to older numerical methods. Two deal with difficult, essentially classical problems in the field of differential equations. The third is concerned with a new and significant algebraic problem, whose successful solution may have a profound effect on military and economic planning, and on administrative procedures affecting the national economy.

These papers state problems; they do not give answers. They are unified by the fact that they all look forward hopefully to a new era in mathematics, in which the impossible will become an everyday occurrence. They are presented here as a challenge to the builders of the new computational science and the manufacturers of its tools.

J. H. Curtiss
Chief, National Applied
Mathematics Laboratories.

1. Some Unsolved Problems in Numerical Analysis [1]

Douglas R. Hartree [2]

Introduction

Numerical analysis is the science and art of carrying out numerical calculations. Work in numerical analysis is not, however, only numerical. There is a substantial amount of development and study of general methods, and that will often be algebraic and analytical. But the purpose of such study must be to lead to a practical numerical process; otherwise, that study may be elegant mathematics but it is not numerical analysis. The emphasis on practicable numerical processes may result in very substantial departures from the point of view of the conventional algebraical or analytical treatments of standard textbooks.

Two examples will illustrate this. Consider first the differential equation

$$y'+2xy=1 \tag{1.1}$$

with the boundary condition $y=0$ when $x=0$. The standard method of solution is given in the first chapter or two of almost any elementary text on differential equations. In this case, application of the method gives

$$y=e^{-x^2}\int_0^x e^{w^2}dw,$$

and in the conventional treatment this would be regarded as a complete and final answer to the problem. But it is not a complete and final answer when you regard the problem as one of numerical process as distinct from an analytical formula. How far it is from a final answer can be seen from this consideration: if it is required to evaluate the integral

$$\int_0^x e^{w^2}\,dw,$$

the easiest way is to solve the equation (1.1) numerically first and then obtain the value of that integral as $e^{x^2}y$.

Another example, also from the field of differential equations, is provided by the equation

$$y''=f(y). \tag{1.2}$$

This also is an elementary form that is included in the first or second chapter of any book on differential equations. The standard method of solution is to multiply by $2y'$ and integrate, obtaining

$$2y'y''=2f(y)y',$$

$$(y')^2=2\int f(y)dy,$$

and finally,

$$x=\int dy/[2\int f(y)dy]^{\frac{1}{2}}. \tag{1.3}$$

As in the first example above, this conventional formal approach is not the best for solution by numerical process. If we try to evaluate the solution from formula (1.3), we have to carry out the quadrature for x in terms of y and then to invert the results to give y in terms of x. All of this is troublesome to carry out numerically. The singularities in the integrand at the ends of the range are very awkward in numerical work, and almost always, if one wishes to solve the equation by numerical process, it is better to take the equation in the original form (1.2). The second-order equation with first derivative absent is the simplest of all kinds of differential equations to solve numerically, and in this example the equation in its original form (1.2) is already of this kind. If it is taken in that form, the process of integration goes quite easily. If the standard textbook method of reducing to quadratures is used, one obtains formulas that are usually numerically more awkward to handle than the original differential equation.

These two examples illustrate that it is often necessary to depart from the conventional textbook treatment of analytical or algebraic problems when the emphasis is on the process for obtaining numerical answers.

It is, of course, rash to talk about "unsolved" problems in numerical analysis; what I really mean, of course, is problems to which I do not know the answers. What I propose to do is to present a series of questions, not concerning large, spectacular problems like the prediction of the weather by numerical integration of the equations of the motion of the atmosphere—which is a possible problem in numerical analysis—but a number of much smaller questions, ones that I have come across in the course of my own work; problems to which the answers should be known, or

[1] Preparation of this paper was sponsored in part by the Office of Naval Research, United States Navy.

[2] Cavendish Laboratory, Cambridge University, England. Acting Chief, NBS Institute for Numerical Analysis, July through October 1948.

at least for which the methods of finding the answers should be known, before the larger problems are tackled. Much of what I have to say is not specifically related to the large digital machines. Numerical analysis certainly will be considerably affected by such machines when they appear, but it is a subject that exists independently of those machines. Much of the experience that has been accumulated to date has necessarily been acquired without those machines, and many of the problems that have arisen are not specifically related to them.

Elimination of approximately known roots of polynomial equations

The first problem to be considered is an elementary one, to which an answer may well be known already, namely, this: Given a polynomial equation in one variable.

$$P(x) \equiv A_0x^n + A_1x^{n-1} + \cdots + A_n = 0$$

and a number p, less than n, of approximate values of roots; let us say $x = \xi_1, \xi_2, \ldots, \xi_p$ are approximate roots, where $p < n$. Then it is desired to eliminate these roots and thus obtain an equation of lower order to be solved for the remaining roots. This is a particularly useful procedure if some of the roots are real because after eliminating the real roots, which are comparatively easy to find, the resulting equation, which has only complex roots and which is of lower, perhaps much lower, order, is easier to solve. The polynomial equation with the roots $\xi_1, \ldots, \xi_p$ is

$$P_p(x) \equiv (x-\xi_1) \ldots (x-\xi_p) = 0.$$

Division of $P(x)$ by $P_p(x)$ gives a quotient, which is a polynomial of degree $n-p$, and a remainder

$$\frac{P(x)}{P_p(x)} = Q(x) + \frac{R(x)}{P_p(x)}.$$

If the roots $\{\xi_i\}$ were accurate, the remainder would be zero. Usually, however, the roots $\{\xi_1\}$ will not be accurate, and, even if they were, there would be rounding-off errors in the coefficients of P_p when the factors are multiplied. Obviously, if the approximate roots $\{\xi_i\}$ were replaced by the correct roots of the equation, different coefficients in the quotient polynomial Q would be obtained. Thus, if it is desired to use that polynomial Q to find the other roots, it is necessary to correct the quotient Q gotten by dividing P by the product obtained from the approximate roots. Hence the problem is first to obtain, from the coefficients in the remainder, corrections to the approximate roots $\{\xi_i\}$ and second—perhaps even more important—to obtain the corrections to the coefficients in the quotient polynomial Q.

Another related problem is how to eliminate a known solution or a set of known solutions from a set of simultaneous, nonlinear equations. Again there may be a known process for handling this problem. But remember that by a process, in this context, is meant one which is amenable to numerical calculation; it is not sufficient to give an algebraic process to ensure that the process suggested is one which can be used in a numerical form. Perhaps one can obtain some guidance from experiments on quite simple examples, say quartics of which two roots are known (because they were set up to have those two roots), which would be helpful in examining numerical methods of eliminating the roots approximately, then finding means of correcting the roots and the quotient polynomial.

Solution of systems of simultaneous nonlinear algebraic equations

Another problem is the solution of algebraic simultaneous, nonlinear equations. The term "algebraic equation" is used here, for want of a better, as an antithesis to "differential equation"; in this sense the equation

$$e^{xy} = x + y$$

is "algebraic". There is one particularly important problem of this kind; although the situation expressed in algebraic terms is independent of the physical situation, the physical situation here is rather significant. In the analysis of crystal structures by means of X-rays, the observed results consist of a number of intensities of reflection of an X-ray beam from different crystal planes. The amplitude of the X-ray beam reflected from the (h,k,l)-plane of the crystal is a sum over positions of atoms in the unit cell of this kind:

$$F(h,k,l) = \sum_j f_j(h,k,l) \cos (hx_j + ky_j + lz_j),$$

where f_j is the scattering factor for a single atom, the j^{th} atom in the unit cell, for the angle of scattering corresponding to reflection from the (h,k,l)-plane, and x_j, y_j, z_j are the coordinates of the atom specified by the suffix j. Now if one could observe the amplitudes F, then one could find the atomic positions by a Fourier transform. Unfortunately, one cannot observe the F's; all that can be observed are their magnitudes $|F|$. Therefore a Fourier transform cannot be used in this way.

One way in which this problem is handled at present is to guess the values of x_j, y_j, z_j by means of all the indications one can obtain from the physical and chemical knowledge of the substance forming the crystal being studied, and to hope that the guess is good enough to determine the signs of the amplitudes F; then these signs are used in the Fourier synthesis to derive the distribution of the scattering electrons. This is equivalent to determining x_j,y_j,z_j. However, as a process of numerical analysis, this method is not satisfactory because it may depend too much on the physical and chemical ideas one has about where the atoms are likely to be in the molecule or in the crystal cell. Also it has the following disadvantage. The measured values of the quantities $|F|$ are subject to experimental error, and the quantities $f_j(h,k,l)$ are not known exactly, so that an exact solution of the equations is not to be expected. It might then be possible to start with a wrong idea of the configuration of the molecules as a whole and obtain a "best" solution on this basis, and this might be accepted as an answer, though really spurious. This might be avoided if one could treat the solution of the equations simply as a problem of numerical analysis and obtain the values x_j,y_j,z_j for a set of equations of that kind without using the step which depends on physical and chemical intuition. In this particular context the values of F and f are observed for a large number of values of h, k, and l,—perhaps even up to some hundreds and in the case of protein crystals even up to some thousands—so there are quite a large number of equations and quite a large number of unknowns.

This is just one example of a general situation in which a solution or solutions of a large number of nonlinear equations

$$f_n(x_1, x_2, \ldots, x_j)=0 \qquad (1.4)$$

are required. One possible way of solving such a set of equations is to form the sums of the squares of the f's:

$$2S(F)=\sum_n f_n^2$$

and find the minima of the function S with respect to $x_1 \ldots x_j$. A minimum may be found by a trial-and-error method in which we take a set of trial values x_1 to x_j and attempt to determine how to modify them in order to get nearer a set which minimizes S. For short, let $\mathbf{x}$ be written for $(x_1, x_2, \ldots, x_j)$. At the point $\mathbf{x}$, let us find the direction of the negative gradient of S, i. e., the direction from $\mathbf{x}$ in which the quantity S decreases as fast as possible; the components of $-(\text{grad } S)$ are given by

$$\frac{\partial S}{\partial x_i}=-\sum_n f_n \frac{\partial f_n}{\partial x_i}.$$

We then modify the trial values of $x_1 \ldots x_j$ in the direction indicated, and repeat the process. In this way we obtain an approximation to a "curve of steepest descent," i. e., a curve at each point of which S decreases as fast as possible, since its tangent at each point is in the direction of $-(\text{grad } S)$ there. Now S is nonnegative, so it cannot decrease indefinitely; and speaking descriptively, if you are always going downhill, you must sooner or later reach the bottom. And the bottom is certainly a minimum of S. There may be minima which do not make S zero, but if a minimum of S is not only a minimum but is also one at which S is zero, then the x_j's arrived at form a solution of equations (1.4). It is interesting to note that this method has been used on a differential analyzer to find solutions of simultaneous equations.

This is a process that can well be studied by what I have called [1]* "experimental arithmetic," and such a study may show features of the process which would hardly be suspected from a purely algebraic study. A simple pair of equations like the following may be tried:

$$xy=3$$

$$2x+3y=8.$$

By using the above process on such a simple system of equations, much may be learned about it and the possible practical difficulties which may arise in applying it to larger systems. Of course, in doing such experimental arithmetic for this purpose one must "play fair." One must not manipulate these equations in any of the ways that their simple form happens to permit but which could not be used on general equations. For example, neither algebraic nor numerical methods of elimination of the variables should be used. It must be kept in mind that this is a small-scale experiment for trying out a general method; one is interested not in the solution of this particular pair of equations, but in carrying out the arithmetic as a simple example of a process which later may be applied on a large scale using an automatic machine.

A further point is that some estimate is needed of the relative numbers of operations involved in these small-scale experiments and in the large-scale work which might be done on an automatic machine when there are 50 or 100 equations of this kind. Whether a method practicable on a small scale will also be practicable on a large scale depends very much on the way in which the number of operations needed increases with the number n of equations; whether it increases as a power of n (as $n^2,n^3,n^4,\ldots$) or exponentially with n (as $4^n,5^n,\ldots$) or perhaps as $2^{2^{2^{2^n}}}$. If n equals 1 or 2 or 3, n^4 and 4^n are not very different; in fact, if n is 2 they are just the same. For large n, it matters very much if the number of operations increases as n^4 or 4^n.

Thus the relative numbers of steps used in two methods of solving two or three equations is no guide to the relative numbers of steps that will be used in the solution of, say, 50 equations unless the way in which the number of steps

*Figures in brackets indicate the literature references at the end of this paper.

increases with n is known. One must be able at least to estimate the relative orders of magnitude of the numbers of operations before it can be decided whether the small-scale numerical experiment will be useful as an indication of the methods that should be used on a large scale.

Since there are only two equations in the particular example under consideration, one can draw the contours of constant S in an (x,y)-plane (fig. 1.1) and this figure can be used to illustrate the general process. If small steps in $\mathbf{x}$ are taken, a good approximation to a curve of steepest descent (shown broken in fig. 1.1) is obtained; but there is no purpose in obtaining such a curve exactly: it is only a means to an end, namely, to a minimum of S, and this might be reached more quickly by some modification of the process. It would be better, for example, not to have to evaluate the direction of $-(\text{grad } S)$ very often because each evaluation of (grad S) involves the calculation of n quantities—its n components—and one would get as much information, and perhaps more, by calculating the values of S itself at n different points. Therefore, one might start from some initial point χ such as P in figure 1.1; determine the direction PQ of $-(\text{grad } S)$ there, i. e., the direction of the inward normal to the surface S=constant through P; proceed in that direction, not changing the direction but calculating a number of points on this normal until you identify an approximate location R of the minimum of S on it; and only then recalculate (grad S).

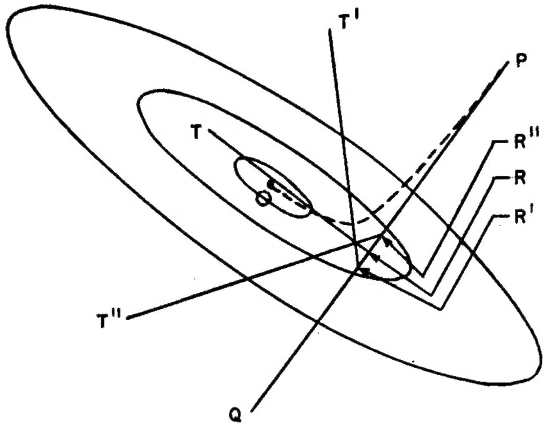

FIGURE 1.1.

In trying this process even on these simple equations, one finds indications of a possible practical difficulty. In the neighborhood of a solution, the contours S=const. will usually be ellipses, and often, perhaps usually, long, narrow ellipses. In such cases, the direction of the normal to the surface S=constant at R, which is going to be the direction in which next to proceed, swings round very rapidly in the neighborhood of the major axis, so that a point has to be determined rather accurately in order to get a good estimate of the direction in which to move next. If it were not well determined, then instead of going in the direction RT (fig. 1.1), you might move in the direction $R'T'$ or $R''T''$. If you like to describe the situation topographically, what you have got is a valley along the major axis OB (fig. 1.2) with sides sloping very steeply upward and the floor of the valley sloping quite slowly upward along OB. Now, so long as you are well away from the floor of that valley, the "steepest descent" process rapidly brings you down into the floor of the valley, but after you get into the floor of the valley the result of approximations in the location of successive points R gives a path bouncing from side to side of the valley, as indicated in figure 1.2, instead of going straight down the middle. What one wants then is some process of damping out that oscillation and making the numbers keep down on the floor of the valley instead of bouncing from side to side.

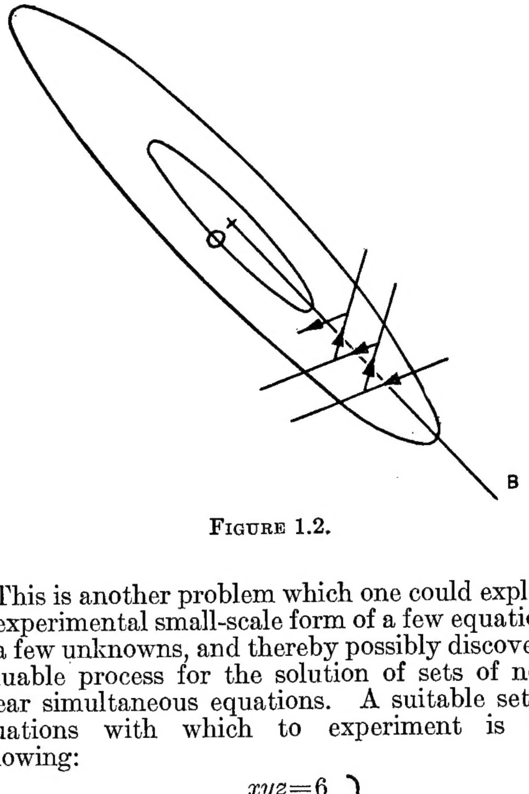

FIGURE 1.2.

This is another problem which one could explore in experimental small-scale form of a few equations in a few unknowns, and thereby possibly discover a valuable process for the solution of sets of nonlinear simultaneous equations. A suitable set of equations with which to experiment is the following:

$$\left.\begin{aligned} xyz&=6\\ x^2-y^2+z^2&=6\\ x+2y+3z&=10 \end{aligned}\right\} \qquad (1.5)$$

Again, remember to "play fair", and do not use any process of algebraic or numerical elimination that could not be used on a general set of equations. Much may be learned in trying numerical work on a simple set of equations such as this.

Two Problems Concerning Relaxation Methods

Two other problems are concerned with the relaxation method of handling differential equations: ordinary differential equations with two-point boundary conditions or partial differential equations with boundary conditions all around the field of integration, such as one is likely to get with equations of elliptic type.

The first of these arises in the theory of the laminar boundary layer in fluid dynamics in connection with the equation [2, 3]

$$y'''+yy''-\beta[(y')^2-1]=0 \tag{1.6}$$

with boundary conditions,

$$y=y'=0,\ x=0 \tag{1.7}$$

$$y'\rightarrow 1,\ x\rightarrow\infty. \tag{1.8}$$

For a positive value of β the solution of this equation is unique. For a negative value of β the solution is not unique. When this nonuniqueness was found in the course of an evaluation of solutions of this equation by means of a differential analyzer, it was suggested, on physical grounds, that the boundary condition to be applied at infinity was not (1.8) but a rather more stringent one, namely, that $y\rightarrow 1$ as fast as possible subject to $y'<1$ for all x.

In the solution of equation (1.6), $y''(0)$ is not specified but has to be determined to satisfy the condition at infinity. If, for a negative value of β, solutions are evaluated with different values of $y''(0)$, there will be one for which y' tends to unity from below as $x\rightarrow\infty$, but does so faster than any other solution.

The situation is represented diagrammatically in figure 1.3; each curve represents a solution of (1.6) with the same value of β but with different values of $y''(0)$. The curve B represents that solution for which $y'\rightarrow 1$ from below and most quickly.

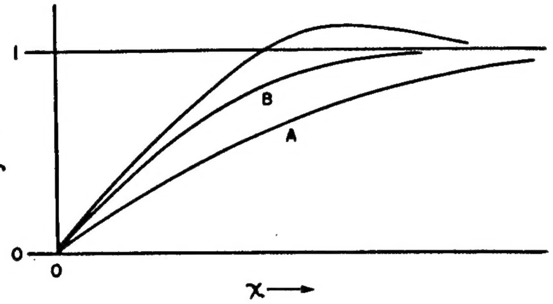

FIGURE 1.3.

The treatment of this equation by relaxation methods has recently been studied by Fox [4], who finds the interesting and curious result that even if one starts the relaxation process from a trial function which is a solution, but not one for which $y'\rightarrow 1$ most quickly, such as the solution represented by curve A in figure 1.3, the process converges to the solution for which $y'\rightarrow 1$ most quickly, just as if it were insisting that the solution you ought to want is precisely the one which has this property. Now why? There is something here that deserves further investigation. It is probably concerned with the stability of the different members of this set of nonunique solutions. But it is the kind of problem to which we ought to know the answer before we start using automatic machines on relaxation or similar methods for equations of this kind.

The other problem also relating to relaxation methods is concerned with their application to a common form of partial differential equations. The usual presentation of the application of relaxation methods to partial differential equations leads one to feel that they cannot be expected to work for equations of hyperbolic type but that they should work for equations of elliptic type. However, during the war, I was concerned with some work on attempting to calculate results such as reflection coefficients at corners in wave guides, and in this context I came across an elliptic equation in conditions in which it seemed intractable by relaxation methods.

A simple example is illustrated in figure 1.4. Here the heavy lines represent boundaries, and it is required to obtain solutions of the equation

$$\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}=-k^2 v, \tag{1.9}$$

with a given value of k, in the region between these boundaries, with $v=0$ on the boundaries. There will be one solution symmetrical about the diagonal PQ and one antisymmetrical.

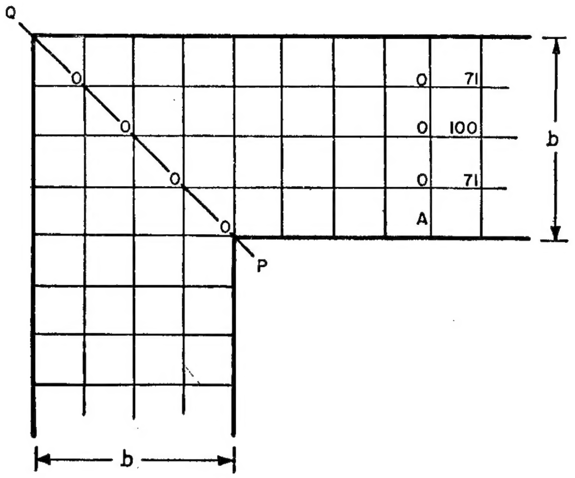

FIGURE 1.4.

If the region in which the solution is required is covered by a square mesh of side a, of which figure 1.5 shows the neighborhood of a typical mesh point, the finite-difference form of equation (1.9) on this mesh is

$$(v_1+v_2+v_3+v_4-4v_0)/a^2=-k^2v_0 \qquad (1.10)$$

or

$$v_1+v_2+v_3+v_4-(4-k^2a^2)\,v_0=0, \qquad (1.11)$$

and, to the approximation represented by the replacement of the partial differential equation (1.10), we have an equation (1.11) for each mesh point and require the solution of this set of simultaneous equations. At first sight it would appear that the equations are very suitable for treatment by relaxation methods. In the case of Laplace's equation, $k=0$ in (1.11); and then there is no difficulty about evaluating a solution with given boundary conditions.

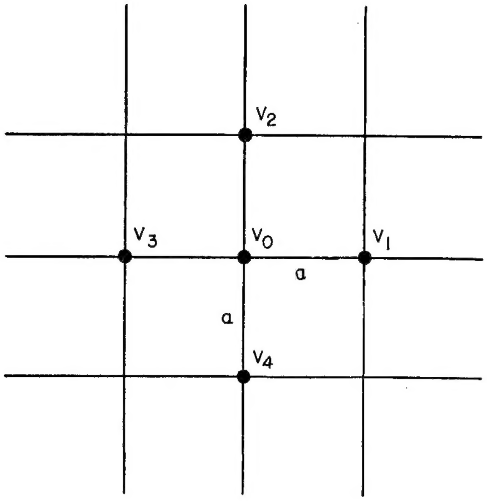

FIGURE 1.5.

Consider first the boundary conditions for the solution that is antisymmetrical about the diagonal PQ. For this solution, $v=0$ on this diagonal, and $v=0$ on the boundaries. Also it is known that far enough from the corner the variation of v across the width of the guide is sinusoidal, so that on the coarse grid (illustrated in fig. 1.3) we can put 71, 100, 71, at the grid points at some selected section.

Now we have an elliptical equation with a closed boundary surrounding the field of iteration, and v given at all points of the boundary, and it looks as if the relaxation process should be an admirable technique to use for its solution in this context. However, this is not so.

That this is not so can be seen in a general way if one considers how the relaxation process might be started. Knowing nothing better, for the first approximation take $v=0$ on the line A and to the left of it. Then the residuals on the line A are positive, and this leads one to insert positive values on line A. Similarly, working to the left from A in figure 1.4, one would try to get a better approximation by filling the whole region with positive final values of v, and there is no indication that you ever have to put in any negative values. On the other hand, it is well known that the variation of v along the length of the guide is ultimately sinusoidal, and therefore you must get negative values somewhere. But the relaxation process gives no indication that any negative values ever have to be introduced anywhere or where to introduce them if at all.

That was bad enough in this case, which is antisymmetrical about the diagonal PQ at the corner for which v has known (zero) values on the diagonal. The situation was much worse for the solution which is symmetrical and for which the values of v on the diagonal are not tied; for this case the process seemed highly unstable. This serves as a warning of the possible difficulties in trying to use this kind of approach, and probably other indirect methods for equations of this kind for which the required solutions have a standing wave character.

Suppose one simplifies the situation by taking the corresponding one-dimensional case. The corresponding finite-difference problem in one dimension is

$$y_{j+1}-(2-k^2a^2)y_j+y_{j-1}=0 \qquad (1.12)$$

which is to be served with y_0, y_n given and not both zero, and k also given. Let

$$\cos\beta=1-\frac{1}{2}\,k^2a^2;$$

then the matrix of the coefficients, with signs arranged to make the diagonal elements all positive, is

$$\begin{bmatrix} 2\cos\beta & -1 & 0 & 0 & \cdots \\ -1 & 2\cos\beta & -1 & 0 & \cdots \\ 0 & -1 & 2\cos\beta & -1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

and its determinant is $[\sin(n+1)\beta]\sin\beta$. This changes sign as n increases from the integer next less than $(\pi/\beta)-1$ to the next higher, so that the equivalent quadratic form is not always positive-definite; and when it is not positive-definite the relaxation process may not be stable.

It is of course possible to replace the equations (1.12) by a system derived from a positive-definite form, but this corresponds, in the two-dimensional case, to using the finite-difference form of the equation $\nabla^4v=k^4v$ instead of that of $\nabla^2v=-k^2v$, and anyone who has used relaxation methods on the biharmonic equation knows that that is a thing to avoid if possible. How to adapt the relaxation process to form a practicable method of dealing with the situation is, as far as I know, one of the unsolved problems of numerical analysis.

Characteristic value problems in ordinary differential equations

Another problem is the best way of using automatic digital calculating machines for evaluating solutions of ordinary differential equations with two-point boundary conditions. In solving ordinary differential equations numerically by almost any process, either on a desk calculator or on a differential analyzer, what really matters is not what the equations are, nor their order, nor degree, but the boundary conditions. And in particular, not what the boundary conditions are but *where* they are. Are they all given at one point of the range of integration or are some given at one point of the range and others at another point of the range? Incidentally, in partial differential equations one gets the same difference between equations for which the boundary conditions completely surround the field of integration and those for which the boundary condition is such that the field of integration is open on one or more directions. That difference is often very much more important than the detailed form of the equations themselves.

In the case of ordinary differential equations with two-point boundary conditions, or with integral conditions on the solution as a whole, there will be certain parameters which have to be determined so that the solution satisfies the conditions specified. These parameters may either be unspecified initial conditions, such as $y''(0)$ in the case of equation (1.6) with boundary conditions (1.7) and (1.8), or constants in the equations themselves. The latter situation occurs, in particular, in characteristic problems in linear differential equations.

Consider first a single homogeneous linear equation with a parameter for which the characteristic values, and corresponding solutions, are to be determined; for example,

$$\frac{d^2y}{dx^2}+[V(x)+A]y=0 \qquad (1.13)$$

with boundary conditions $y=0$ at $x=0$, a, and perhaps a condition to insure that the trivial solution $y\equiv 0$ is not obtained, such as

$$\int_0^a y^2\,dx=1. \qquad (1.14)$$

Now I do not know the best way of programing that situation for automatic machines. One way of handling it—the way often adopted in work with desk machines—is simply to try a number of different solutions, satisfying the boundary condition at $x=0$, with trial values of A until one obtains the solution which satisfies the condition at the far end $x=a$ of the range of integration. That is one possibility, but it means doing a great many computations that have no value except as intermediate steps toward obtaining the solution desired. I have a very definite feeling that that is not the best way to try to do it on automatic machines, but I do not have a method which I regard as satisfactory to put in the place of it.

In this case, when there is only one equation, the problem of normalizing the solution, that is, of satisfying the condition (1.14), can be handled after a solution satisfying the boundary conditions has been obtained, by multiplying this solution through by a constant factor. But consider the corresponding problem with two simultaneous equations:

$$\left.\begin{aligned}\frac{d^2y}{dx^2}+[f_1(x)+A_1]y+[g(x)+B]z=0\\ \frac{d^2z}{dx^2}+[f_2(x)+A_2]z+[g(x)+B]y=0\end{aligned}\right\} \qquad (1.15)$$

with conditions

$$y=z=0 \text{ at } x=0, a \qquad (1.16)$$

and

$$\int_0^a y^2\,dx=\int_0^a z^2\,dz=1; \qquad (1.17)$$

$$\int_0^a yz\,dx=0. \qquad (1.18)$$

The normalization cannot now be postponed until after the solution satisfying the conditions (1.16) is obtained. It must be done in the course of determining the solution. For given values of $y'(0)$, $z'(0)$ there are two parameters, A_1 and A_2, adjustable to obtain solutions that behave themselves as you want at both ends of the range of integration. And there are the values of $y'(0)$ and $z'(0)$ to adjust in order to satisfy the normalizing conditions (1.17). The value of B must be adjusted so that the solutions satisfying the other conditions also satisfy the orthogonal condition (1.18). This is a simple example of the situation that arises in trying to evaluate atomic structures. But in this context there is a further complication in that the functions f_1 and f_2 and g are themselves related in a nonlinear way to the solutions y and z of equations (1.15) so that there are two equations for y and z and three other equations, to specify the way f_1, f_2, and g depend on the solution y and z. With this degree of complication, I know at present of no other way of working than by trial. One method would be to start with estimates of the functions y and z; use these in the equations which determine the functions f_1, f_2, and g; then evaluate the solutions y and z of equations (1.15) with these functions for f_1, f_2, and g; and repeat the process, taking these solutions of (1.15) as better estimates, until a final result is obtained in which this process reproduces the estimated y and z functions. Such an iterative process, however, does not always converge, and a better method of improving the estimates of the y and z functions must be used. This is another class of problems in which procedures and programs for solution on automatic digital machines are very much needed.

Motion of a continuous distribution of charge under mutual forces between its parts

In many if not all cases the solution of a set of equations can be handled without knowing the physical situation to which they refer. However, in many cases knowledge of the physical context of the equations is helpful both in discussing them and in suggesting means for their treatment. This is true, for instance, in the situation where a distribution of electrical charge, which can be treated as continuous, is moving under the influence of the mutual forces between its parts so that the field acting on any one charge depends on the distribution on all other charges, which themselves are so moving. This involves two kinds of equations: the equations of motion of the individual charges and Poisson's equation for the field arising from the space-charge distribution.

There may be some rather curious boundary conditions in the solution of Poisson's equation. Usually, the boundary condition is the value of the potential at the boundary and nothing is given about the potential gradient. However, if a bounding surface on which the potential is given can emit electrons and the velocity distribution of the electrons can be neglected, then there is an additional relation at the surface, namely, one between the emission current and the potential gradient. If the potential gradient is positive, there is no emission; and if the potential gradient is negative, there is the full temperature limited emission. Intermediately there is a condition in which the potential gradient at the surface is zero, and the emission current adjusts itself to give just this value of the potential gradient at the surface (space-charge limited emission). This gives rise to a rather unusual situation when you try to determine the solution of Poisson's equation because it is necessary to combine the problems of determining the solution of Poisson's equation and the emission current distribution.

Some psychological problems in numerical analysis

There are several psychological problems in numerical analysis, some of which are directly relevant to the use of automatic digital machines and must be borne in mind when programing a problem.

One of the unsolved problems of numerical analysis is how to overcome the attitude of the mathematical fraternity toward the subject—an attitude exemplified by the comment of a distinguished mathematician, introducing a lecture of mine on the mechanical integration of differential equations, that he had always regarded the solution of differential equations as "a very sordid subject."

Another problem is that of getting what I have called a "machine's eye-view" of a problem as presented to an automatic machine. It must be remembered that the machine will carry out the instructions given to it literally and blindly with no exercise of intelligence beyond these instructions. In doing a numerical calculation by hand process, one uses one's own intelligence, almost unconsciously, very much more than one really realizes. For example, suppose that in the course of a calculation a division has to be performed where the divisor turns out to be zero. In such a situation there are many things which a human computer might do. He might just knock off and go home to lunch and give himself time to think out what had gone wrong. What he would certainly *not* do would be to go on forever trying to divide by zero; but that is precisely what an automatic machine will do unless it is specifically told not to by instructions deliberately included in its program for the purpose. Therefore, in programing a problem, all the unusual situations that might arise in the course of the solution of the problem must be anticipated, and the machine must be given adequate instructions to identify each one and to take the appropriate action if any one or any combination of them occurs. They probably will not occur, but the instructions provided to the machine must prepare it for whatever may happen. Abstracting oneself from the intelligence that the human computer applies in the course of a calculation is a good deal harder than one realizes until one tries to do it. But it is important that one should try and cultivate what I call the "machine's-eye view" of the sequence of instructions that go into a machine.

A third psychological problem is the problem of getting enough "feel" for how the calculation is going when it is being done by an automatic machine. If one is actually handling the numbers oneself, one has a feeling for how the work is going which is difficult to get from seeing the completed results of the work of someone else and which seems almost impossible to get if the mechanism does the details of the work and never even exhibits them. My own experience with the differential analyzer has been that even a solution on this machine is too automatic to permit one to get a real feel for the way the calculation is going; on a problem of a new kind it has almost always been worth while to carry out the evaluation of one solution by hand myself to get a feel for the relative magnitudes of the variables, and the way in which they behave, before turning it over to the machine. In simple cases it might be possible, and advisable, to examine the problem analytically before doing numerical work. But in more complicated cases the analytical treatment may be too difficult or long, and one would have to depend on the numerical results themselves to give one a feeling for the general way in which they are behaving and their general character as distinct from detailed numerical results for special cases. It is the problem of getting that feeling and intuition for the way calculations are going when the details of the calculations are carried out by an automatic machine which, I think, is the third of these psychological problems of numerical analysis.

References

[1] D. R. Hartree, Experimental arithmetic, Eureka **10,** 13 (1942).

[2] V. M. Falkner and S. W. Skan, Solutions of the boundary layer equations, Phil. Mag. [7] **12,** 865 to 896 (1931).

[3] D. R. Hartree, On an equation occurring in Falkner and Skan's approximate treatment of the equations of the boundary layer, Proc. Cambridge Phil. Soc. **33,** 223 to 239 (1937).

[4] L. Fox, The solution by relaxation methods of ordinary differential equations, Proc. Cambridge Phil. Soc. **45,** 50 to 68 (1949).

2. Numerical Calculations in Nonlinear Mechanics[1]

S. Lefschetz[2]

Anyone scanning through the Minorsky report [1]* on nonlinear mechanics will convince himself that as a rule, on this subject, mathematics is only able to provide qualitative and very imperfect information. This is where the numerical analyst may step in and complete the work. In particular, here, as in so may other cases, the computing machines may often guide mathematical research by providing information on "the proper direction of motion" for the investigator. The following two simple examples will illustrate what we have in mind.

The first example is the famous equation of van der Pol, which arises very naturally in the following manner. Consider first the standard so-called LRC equation

$$L\dot{x}+Rx+\frac{1}{C}\int xdt=E \qquad (2.1)$$

for an electric circuit (fig. 2.1) with current x, constant emf E, inductance L, resistance R, and capacitance C. By differentiating with respect to t, the relation (2.1) is replaced by

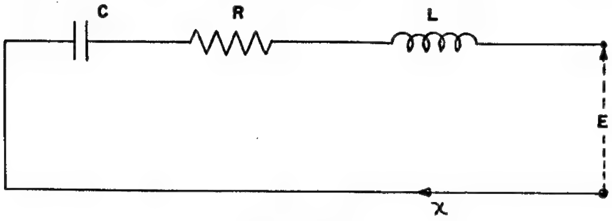

FIGURE 2.1.

$$L\ddot{x}+R\dot{x}+\frac{1}{C}\,x=0, \qquad (2.2)$$

which is a linear homogeneous differential equation with constant coefficients. The explicit solution of (2.2) offers no difficulty and is found in every sophomore calculus text.

In equation (2.1) the middle term Rx represents a dissipation proportional to the current x. This is the simplest assumption possible and well in agreement with observation for ordinary resistors and moderate variations of temperature. Under less simple circumstances one may have to replace Rx by more complicated expressions. This is notably the case when the circuit (fig. 2.2) contains a vacuum tube. Neglecting certain things and in particular the grid current, it turns out that in this case, when x is very small, the circuit behaves as if it contained a negative resistance. The simplest assumption that one can make, compatible with this fact and still having a dissipative term which is an odd function of x, is to take for this term a cubic polynomial $\alpha x^3/3-\beta x$, α and β positive. Instead of (2.2) we have now the equation

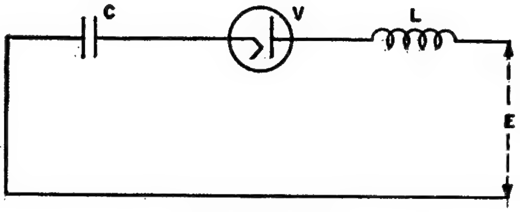

FIGURE 2.2.

$$L\ddot{x}+(\alpha x^2-\beta)\dot{x}+\frac{1}{C}\,x=0. \qquad (2.3)$$

By the familiar method of selecting suitable time and current units, (2.3) may be put in the "dimensionless" form

$$\ddot{x}+\mu(x^2-1)\dot{x}+x=0, \qquad (2.4)$$

which is van der Pol's equation. In this equation the only variable parameter is μ and it must be positive. By a familiar method, (2.4) is often replaced by a pair of equations of the first order of a type investigated at length about three quarters of a century ago by Poincaré:

$$\dot{x}=y,\ \dot{y}=-x+\mu(1-x^2)y. \qquad (2.5)$$

The solutions of (2.5) are pairs of functions of time $x(t)$, $y(t)$, which represent graphs covering the "phase-plane" xy.

In systems such as (2.5) there are, according to Poincaré, two all important elements: the critical points and the solutions which are closed curves or limit-cycles in Poincaré's terminology. The critical points correspond to the positions of equilibrium or "static" steady states of the associated physical system, and the limit-cycles to its oscillatory steady states. Both may be stable or unstable, and it is usually the stable steady states that are important. A very remarkable result due to Poincaré asserts that whatever the initial conditions of the physical system it will gravitate toward one of the steady states. It may thus acquire spontaneously a steady oscillation, referred to in general as a self-oscillation. This will not occur in the initial linear circuit with ohmic resistance. These self-oscillations turn out to be highly important in applications.

[1] The preparation of this paper was sponsored by the Office of Naval Research, United States Navy, Contract N6ori-105.

[2] Princeton University, Princeton, N. J.

*Figures in brackets indicate the literature reference at the end of this paper.

Now the amount of exact mathematical information available regarding van der Pol's equation (2.4) or the associated system (2.5) is rather meager and may be summarized as follows:

(a) There is only one state of equilibrium, the origin $x=0$, $y=0$, in the phase plane (inactive circuit), and it is unstable.

(b) There is exactly one limit-cycle $C(\mu)$ in the phase-plane; it is a unique oscillation for each μ. The limit-cycle surrounds the origin. The existence of this unique oscillation was established about 20 years ago by the French physicist Liénard.

(c) For μ very small $C(\mu)$ is very close to the circle of radius 2, and the oscillation is practically the barmonic oscillation $x=2\sin t$, $y=2\cos t$. Here of course the time origin has been chosen to have zero phase.

(d) The limiting position of the limit-cycle for μ very large is also known and there is information about the order of magnitude and the frequency of the oscillations for large μ [2]. These oscillations for large μ are of the well known relaxational type, a term introduced in this very connection by van der Pol.

Properties (a) and (b) together imply that whatever the initial conditions the van der Pol system tends to become oscillatory and the oscillation, represented by the limit-cycle, is absolutely well defined.

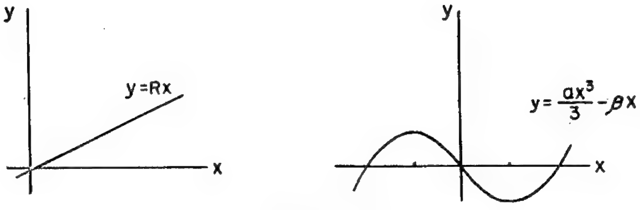

FIGURE 2.3.

To sum up, then: There is ample information for μ very large or very small but nothing in between. This is where numerical analysis might step in—to provide as ample information as desired on the amplitude $a(\mu)$ and the frequency $f(\mu)$ of the oscillations as functions of μ. It is interesting to point out that van der Pol obtained values of these functions for $\mu=0.1, 1, 10$ by means of the graphical method of isoclines. That is to say, for each of these three values of μ he plotted a fairly large collection of solutions in the phase plane, observed the presence of a unique oscillation in each case, and made reasonable estimates of what goes on in general. His graphs have been published in many articles of his and others on these questions.

The situation becomes far more involved when the impressed emf in figure 2.2, instead of being constant, is itself oscillatory. In the simplest case—a simple sine wave—(2.1) becomes

$$L\dot{x}+Rx+\frac{1}{C}\int xdt=E\sin(\omega t+\alpha),$$

which yields instead of (2.4), the relation

$$\ddot{x}+\mu(x^2-1)\dot{x}+x=k\mu\omega\cos(\omega t+\alpha) \qquad (2.6)$$

much investigated by Littlewood [3], Cartwright [3, 4], and by Levinson [5]. The important practical question is whether subharmonic resonance takes place—that is to say, whether there exist oscillatory solutions $x(t)$ of (2.6) whose frequency is a fraction of the frequency $\omega/2\pi=f$ of the impressed oscillation. The authors in question have obtained some general qualitative information on this question. Much more accurate information may certainly be obtained by numerical analysis.

Returning again to van der Pol's equation in the form (2.4): There is considerable evidence that the assumption that Rx in (2.1) is to be replaced by a mere cubic is too simple. It is likely that in practice a higher degree polynomial, and one not necessarily odd, may be necessary, or conceivably some other function. We would then still have ample qualitative information, but would certainly find good use for modern computational methods to obtain accurate quantitative information.

Before leaving van der Pol's equation it may be pointed out that Rayleigh's equation for the rectilinear motion of a mass particle under friction [6]

$$m\ddot{y}+(B\dot{y}^2-A)\dot{y}+ky=0, \qquad (2.7)$$

A and B positive, is reducible to the van der Pol form by adopting a new variable $\alpha x=\dot{y}$, and making a suitable change of time scale. Thus van der Pol's equation is of practical interest even outside of circuit questions. However, in problems of the nature here considered, circuits are very convenient, in that they enable one to imitate most economically in the laboratory physical systems of very great complication.

Whatever the reasons, consider another electrical problem on which a good deal has been written by physicists and electrical engineers: the phenomenon known as ferroresonance [7,8,9]. Let the inductance in an LRC circuit consist of a coil with iron core and let there be impressed a sinusoidal emf, $E\sin\omega t$. Let it be assumed that the resulting current is periodic with the same period as the voltage and has a Fourier series representation

$$x=I_1\sin(\omega t+\alpha_1)+I_3\sin(3\omega t+\alpha_3)+\cdot\cdot\cdot\cdot .$$

The effective emf is $e=E/\sqrt{2}$, and the effective current is

$$i=\frac{(I_1^2+I_3^2+\cdot\cdot\cdot)^{1/2}}{\sqrt{2}}.$$

If one plots i against e there is obtained a curve such as the one in figure 2.4. The observed facts

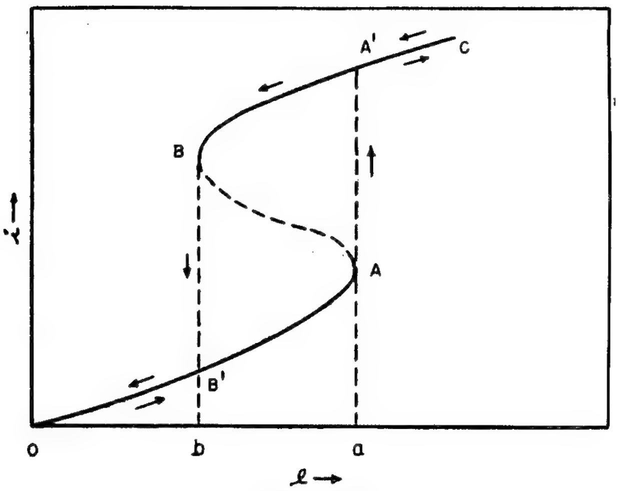

FIGURE 2.4.

are as follows: as e increases from zero, the effective current $i(e)$ is given by the ordinates of the arc OA; when e passes the value a, $i(e)$ jumps suddenly from the value aA to the larger value aA', and follows then the upper arc $A'C$. Similarly as e decreases from some position beyond a, $i(e)$ is given by the ordinate of the upper arc until e reaches b when $i(e)$ drops suddenly from the value bB to the value bB'. If e decreases still further below the value b, $i(e)$ is given by the ordinates of the lower arc OB'. Of course this situation arises only when the core reaches saturation. Well below saturation the only significant arc is OA alone and there is no ferroresonance.

The differential equations of the system are obtained as follows: Denoting by ϕ the flux through the coil we will have

$$\frac{d\phi}{dt}+Rx+\frac{1}{C}\int xdt=E\sin\omega t. \qquad (2.8)$$

The saturation curve of the coil yields a relation $\phi=f(x)$, where $f(x)$ is odd and may be approximated by a polynomial of odd degree, at least when the "hysteresis loop" is thin. Upon substituting in (2.8) and differentiating, there results the differential equation

$$f'(x)\cdot\ddot{x}+(R+f''(x)\dot{x})\dot{x}+\frac{1}{C}x=E\omega\cos\omega t, \qquad (2.9)$$

where $f'(x)$, $f''(x)$ are the first and second derivatives with respect to x. Needless to say this equation is far more nonlinear than van der Pol's. About all that one can do with it mathematically is to prove the existence of a solution of frequency $\tau=\omega/2\pi$, or fractions of f [10, 11, 12] (subharmonic resonance), and to calculate the effective fundamental harmonic $i_1(e)$. In point of fact, its graph is then found to be not too far—not more than 10 percent—from the graph of $i(e)$, and to have more or less the same general form [7]. It is clear that there is here again a wide open field for numerical analysis. Some work along this line has already been done. However, much more could certainly be accomplished if the numerical analysis were to be paralleled by a theoretical investigation similar to what has been done for the van de Pol equation.

We may conclude with a very general remark. In the past in dealing with nonlinear questions the tendency has always been to "linearize" the problem or to keep the situation within "linear" bounds. This is, for instance, the source of the vector diagrams encountered in the theory of alternating-current machinery. One endeavors not to reach saturation and hopes for the best. It may now be possible, with the higher technique of modern numerical analysis, to deal directly with the nonlinearized situations, thus coming much closer to understanding what actually goes on.

[1] N. Minorsky, Introduction to non-linear mechanics, Report No. 534, 546, 558, David W. Taylor model basin, Dept. of the Navy, Washington, D. C. (1944–46). This work contains a very ample bibliography, in particular on van der Pol's equation and the work of Poincaré referred to in the text. An excellent summary of early work on van der Pol's equation is: Balth. van der Pol, The non-linear theory of electric oscillations, Proc. Inst. Radio Eng. **22,** 1051 to 1086 (1934).

[2] D. A. Flanders and J. J. Stoker, The limit case of relaxation oscillations. Studies in non-linear vibration theory, p. 50 to 64, Institute for Mathematics and Mechanics, New York University (1946).

[3] Mary L. Cartwright and J. E. Littlewood, On non-linear differential equations of the second order: I. The equation $\ddot{y}-k(1-y^2)\dot{y}+y=b\lambda k\cos(\lambda t+\alpha)$, k large, J. London Math. Soc. **20,** 180 to 189 (1945).

[4] Mary L. Cartwright, Forced oscillation in nearly sinusoidal systems, J. Inst. Elec. Eng. **95** (part III), 88 to 96 (1948).

[5] N. Levinson, A simple second order differential equation with singular motions, Proc. Nat. Acad. Sci. **34,** 13 to 15 (1948).

[6] Lord Rayleigh, On maintained vibrations, Phil. Mag. [5], **15,** 229 to 235 (1883).

[7] W. H. Surber, Jr., A study of ferroresonant and subharmonic oscillations (thesis for the degree of Electrical Engineer), Princeton University (1943).

[8] C. G. Suits, Studies in non-linear circuits, Trans. Amer. Inst. Elec. Engrs. **50,** 724 to 736 (1931).

[9] P. H. Odessey and E. Weber, Critical conditions in ferroresonance, Trans. Amer. Inst. Elec. Engrs. **57,** 444 to 452 (1938).

[10] J. D. McCrumm, An experimental investigation of subharmonic currents, Trans. Amer. Inst. Elec. Engrs. **60,** 533 to 540 (1941).

[11] I. Travis and C. N. Weygandt, Subharmonics in circuits containing iron-cored reactors. Trans. Amer. Inst. Elec. Engrs. **57,** 423 to 430 (1938).

[12] I. Travis, Subharmonics in circuits containing iron-cored inductors, Part 2, Trans. Amer. Inst. Elec. Engrs. **58,** 735 to 742 (1939).

Recent references on forced oscillations: J. LaSalle, Relaxation oscillations. Quart. Appl. Math. **7,** 1 to 19 (1949). S. Lefschetz, Contributions to the theory of nonlinear oscillations, Ann. of Math. Study No. 20 (1950), notably the papers by M. L. Cartwright, J. G. Wendel, and C. E. Langenhop-A. B. Farnell.

3. Wave Propagation in Hydrodynamics and Electrodynamics[1]

Bernard Friedman[2]

There are a number of problems concerning wave propagation which have been formulated mathematically and for which treatment by numerical methods has been suggested. However, these problems still require a good deal of work before they will be ready for solution with high-powered computing machines. A few such problems in applied mathematics will illustrate some of the typical difficulties that require further study.

The difficulties are due to one or more of these four factors: (1) wave propagation in an inhomogeneous medium, (2) existence of complicated boundary conditions, (3) nonlinearity of the appropriate equations, (4) the prescription of boundary conditions on a free boundary and therefore on an unknown boundary. Each of these four difficulties may be illustrated by a practical problem which is still unsolved.

The first factor, wave propagation in an inhomogeneous medium, is vitally important in geophysics, in acoustics, and in electrodynamics. Consider one problem from electrodynamics: the transmission of electromagnetic waves through the atmosphere. Suppose an antenna is placed on some point above the earth's surface. How far do the radio waves travel? For the short wavelengths under consideration, the Heaviside layer has no effect and, by analogy with optics, the rays should travel in a straight line until they reach the horizon. However, a study of the effect of the atmosphere reveals very different results. The variation of temperature with height and the change in the amount of water vapor in the air from point to point affect the index of refraction and therefore the speed of the electromagnetic waves. Because of this, the paths of the radio waves are no longer straight but become curved. In some cases the rays may be trapped, that is, the waves are refracted back so that all the energy is concentrated in a region a short distance above the earth. Consequently, a larger range may be obtained than under ordinary atmospheric conditions. An amusing illustration of this phenomenon occurred last year. The police short-wave system in a Florida town was interrupted by mysterious calls in French. After some investigation it was found that police calls from a Montreal station were being received. The only explanation for such a large range is the atmospheric refraction of the radio waves.

The practical problem is this: Given an antenna (fig. 3.1) emitting radio waves of a fixed frequency and the surrounding atmospheric conditions, determine the range of those waves. The mathematical formulation of the problem is very simple. Suppose the antenna (fig. 3.2) placed at a point S at a height h above the surface of the earth is sending out radio waves of frequency f. The problem is to find a function ψ that satisfies the wave equation:

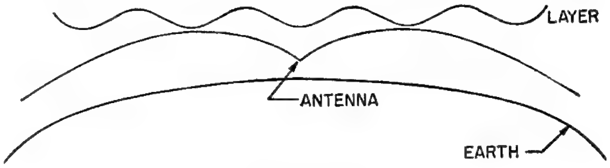

Figure 3.1.

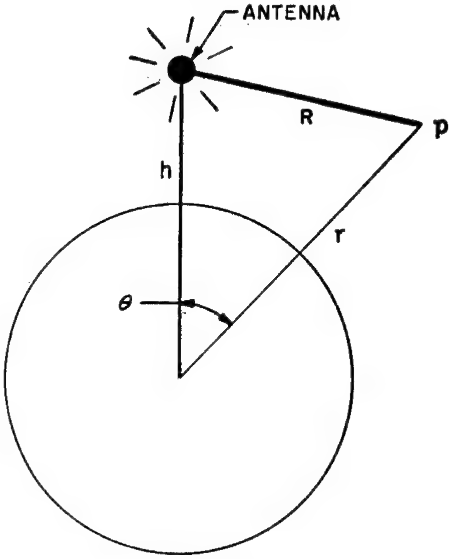

Figure 3.2.

$$\nabla^2\psi - n^2k^2\frac{\partial^2\psi}{\partial t^2} = 0, \tag{3.1}$$

where $k=2\pi f/c$, c is the velocity of light, and n the index of refraction of air. This equation should be solved subject to three conditions: (1) In the neighborhood of S,

$$\psi \sim \frac{e^{-ikR}}{R}, \tag{3.2}$$

where R is the distance from the source S. (2) $\psi=0$ at the surface of the earth This implies that the

[1] Preparation of this paper was sponsored by the Office of Naval Research, United States Navy.

[2] Institute for Applied Mathematics and Mechanics, New York University, New York City, N. Y.

earth is a perfect conductor. Of course, this is not exactly correct, but it is accurate enough for most practical purposes. (3) At infinity ψ must behave like an outgoing wave:

$$\psi\sim e^{-ikr}/r \tag{3.4}$$

where r represents the distance from the center of the earth. Thus the mathematical problem is formulated.

In the solution, the variation of the index of refraction is the quantity that causes all the difficulty. We know that n will vary from point to point, depending upon the temperature of the air, upon the partial pressure of water vapor, and also upon the atmospheric pressure. The question is, given the meteorological conditions as represented by the value of n from point to point, how can this problem be solved? Of course, in this general formulation, the solution would be possible only on a large computing machine. For many purposes, however, it would be sufficient to solve the problem if it is assumed that n varies as a function of r, the distance from the center of the earth; that is, if it is assumed that the atmosphere is arranged in spherical layers around the earth. This is so because under normal conditions the variation in atmospheric change due to height is much more important than the variations due to change in horizontal distances. If spherical polar coordinates with polar axis passing through S are used, we can write a solution for this problem:

$$\psi=\sum_0^\infty c_jP_j\,(\cos\,\theta)U_j(r), \tag{3.5}$$

where $P_j(\cos\ \theta)$ is the j^{th} Legendre polynomial, θ is the angle from the pole, c_j is a constant, and $U_j(r)$ satisfies the differential equation:

$$\frac{d^2U_j}{dr^2}+\frac{2}{r}\frac{dU_j}{dr}+\left(k^2n^2-\frac{j(j+1)^2}{r^2}\right)U_j=0. \tag{3.6}$$

$U_j(r)$ must satisfy certain conditions which can be obtained from the previously given conditions on ψ, (3.3) and (3.4). The difficulty lies not only in solving this equation for arbitrary values of n, but also in evaluating the sum for ψ. In the case of n equal to a constant, this equation can be solved: U is then the Bessel function of order $j+1/2$. To solve the problem for the constant atmosphere, a series of the following kind has to be evaluated:

$$\psi=\sum_0^\infty c_jP_j\,(\cos\,\theta)\ J_{j+\frac{1}{2}}(ka), \tag{3.7}$$

where a is the radius of the earth, and k is as defined previously. Now, this series converges so slowly that more than a million terms would be required in order to get any practical results. The reason for this is that, for 10-cm waves, the number ka is of the order of 5×10^8. For such large values of the argument, the asymptotic form of $J_{j+\frac{1}{2}}\ (ka)$ is the following:

$$J_{j+\frac{1}{2}}(ka)=\sqrt{\frac{2}{\pi ka}}\cos\,(ka-(j+1)\pi/2)$$

as long as j is smaller than ka. Then (3.7) becomes a sum of cosines, which does not converge rapidly because the successive terms decrease too slowly. In order to obtain rapidly decreasing terms, j must be larger than ka. Of course, it would be possible for a large computing machine to handle the summation of a million terms or more, but it should be noticed that these terms oscillate very rapidly and very irregularly, so that to get any sort of accuracy, one would need to start with a large number of digits in each term to avoid the accumulation of errors.

However, this particular problem need not be solved by machine. G. N. Watson, in a brilliant paper [1]* in 1919, showed how this sum could be expressed as a complex integral, and then, by use of a difficult analysis of the Bessel functions, he was able to evaluate the integral as a sum of residues. The difficulties of the analysis are such, however, that it seems hopeless to expect a result similar to that of Watson in the case of general values of n.

During the war a great deal of attention was focused on this problem because of its importance for the Armed Services. After all, if the radar set does not "see" above a certain height, you have a blind spot, and airplanes would be out of view even though the radar is working perfectly. Of course, conversely, under certain atmospheric circumstances, if the pilot knows that the radar has a blind spot, he would try to take advantage of it.

An advance in this problem was made by M. H. L. Pryce [2] in England, who introduced a transformation whereby the surface of the earth became a plane, and the rays were curved. As a result the problem was transformed so that a somewhat simpler equation had to be solved. A modified index of refraction, N, is introduced as follows:

$$N=\frac{r}{a}\frac{n(r)}{n(a)} \tag{3.8}$$

and then a function in rectangular coordinates, $\phi(x,y,z,t)$, must be found, which satisfies this equation:

$$\Delta^2\phi-k^2N^2\frac{\partial^2\phi}{\partial t^2}=0\cdot \tag{3.9}$$

The function (3.5) behaves like a source in the neighborhood of the antenna; it vanishes on the earth's surface ($z=0$); and at infinity it behaves like an outgoing wave. Next, Professor Furry [3] of

*Figures in brackets indicate the literature references at the end of this paper.

Harvard showed that the solution of this partial differential equation depended upon the solution of the following eigenvalue problem: Given N, find the eigenvalues λ such that the solutions U (3) of the equation.

$$\frac{d^2U}{dz^2}+(k^2N^2-\lambda)U=0, \qquad (3.10)$$

satisfy the following boundary conditions: $U=0$ at $z=0$ and U is an outgoing wave at infinity. Here, it is assumed that N is a function only of the distance above the surface of the earth.

Unfortunately, the solutions of this equation are known only for a few functions $N(z)$. One such case is that of a constant atmosphere where $N(z)^2$ becomes a linear function of z and the solutions are Hankel's function of order 1/3. Surprisingly enough, these are identically the same functions that appear in Watson's exact solution of the constant atmosphere case. This equation (3.10) happens to be an approximation to equation (3.6), which seems to be sufficiently correct for the problems being treated here. It is important to notice that in this equation the eigenvalues are complex. The reason is that the boundary condition at infinity is not a self-adjoint one, since it distinguishes between $+i$ and $-i$.

We have been able to go far enough in the question of atmospheric propagation to reduce the problem to an ordinary differential equation of the second order. Now, what is desired is a solution of (3.10) when $N(z)$ is given as an arbitrary function of z, the actual function to be determined by the atmospheric conditions. This would seem to be a simple computing task; yet it turns out to be extraordinarily difficult. First, since even in the simplest case the eigenvalues are complex, we have to deal with both complex values of λ and complex eigenfunctions $U(z)$. Second, the boundary condition at infinity is difficult to handle. How can we recognize that a function is behaving like an outgoing wave, i. e., like e^{-ikz} and not like $Ae^{-ikz}+Be^{ikz}$, where A and B are constants? Unfortunately, if we start with a value of λ that is not the exact one, we must always get a solution containing incoming waves. The only time a solution without incoming waves can be obtained is when the exact value of λ is used, i. e., exact to an infinite number of decimals. Third, since the solutions oscillate very rapidly, we must work with many decimal places in the beginning in order to come out with a few at the end. Because of these difficulties, very little numerical work has been done even though there is a great need for it. The engineer's work would be simplified if this problem could be handled even for functions $N(z)$ which are composed solely of straight-line segments.

This differential equation offers an opportunity for mathematical investigation. We have here an eigenvalue problem which is not self-adjoint. What can be said about the eigenfunctions? Are they complete? Are they orthogonal? Can an arbitrary function be expanded in terms of these eigenfunctions? All the results of the Sturm-Liouville theory are open to generalizations, but so far the work has been negligible, Yet the physicist, in his work in acoustics, electrodynamics and quantum theory, needs the answers to the preceding questions. Every time he considers wave propagation outside of a finite region, he introduces a condition that the function behave like an outgoing wave at infinity.

Consider next a second type of problem which requires mathematical treatment. The difficulty here is in the complicated nature of the boundary. A practical illustration is the problem of the tides.

Tides are forced oscillations of the ocean which result from the periodic disturbances of the earth's surface due to the varying gravitational attraction of the sun and moon produced by relative motion of these bodies. Mathematically we have a linear partial differential equation of hyperbolic type in two space variables and one time variable. As an illustration of the kind of equation, consider the following case. Assume symmetry on the polar axis and suppose the disturbing force varies as $e^{i\sigma t}$. Then we must solve the equation [4]

$$\frac{\partial}{\partial\mu}\left(\frac{h(1-\mu^2)}{f^2-\mu^2}\,\frac{\partial\xi'}{\partial\mu}\right)-\frac{4\omega^2a^2\xi'}{g}=\frac{-4\omega^2a^2\bar{\xi}}{g} \qquad (3.11)$$

$$f=\frac{\sigma}{2}$$

$$\xi=\xi-\bar{\xi},$$

where $\mu=\cos\,\theta$, the cosine of the colatitude; h represents the depth of the ocean, which is not assumed constant; $\omega=2\pi$ times the frequency of the disturbance of the earth's surface; a=radius of the earth; g=the acceleration of gravity; $f=\sigma/2\omega\xi'$. The quantity we are solving for is $\xi-\bar{\xi}$, where ξ is the elevation of the ocean and the height of the tide, $\xi-\bar{\xi}$ is the apparent elevation produced by the disturbing force. In practice, the mathematical theory is used only to a very slight degree. It is assumed that tidal motions have gone on so long that all the free oscillations have been damped out so that only forced oscillations are present. The period of these forced oscillations is known from astronomy. Hence, a trigonometrical series having these periods with arbitrary amplitudes and phases can be set up, and this series can be fitted to the observed tides at any one particular point. After the amplitudes and phases are determined, this experimental series is used to extrapolate the tides at any given time. Thus, it is clear that the dynamic theory is not used at all in the computation of the tides.

Of course, there is considerable literature giving solutions of simpler equations for the tides—equations obtained by neglecting the Coriolis forces due to the rotation of the earth or by making unrealistic assumptions about the shape of the boundary or the shapes of the oceans. The reason for making these simplifying assumptions is a natural one. We try to achieve a formulation which permits us to represent the solution in terms of tabulated functions. But here there seems to be a good opportunity to make use of modern calculating machines to handle this spectacular problem of solving the dynamical equations directly. In this way it would be possible to obtain tides for points of the earth where the previously mentioned type of analysis has not been carried out. Also the tidal waves in mid-ocean could be studied. Questions unsolved to date, such as whether the tides in the open ocean behave like progressive waves or stationary waves, might be answered.

Such a practical program for calculating tidal waves should be feasible since the differential equations are linear and are accurately known. The difficulty will appear at the boundary. Not only must the actual coast line of the ocean be approximated, but also the boundary conditions on it must be determined, since it is not known how much of the energy of the incoming wave is absorbed at the shore and how much is reflected back into the ocean. The extremes of perfect reflection and total absorption could be tried and the results compared with the actual observations. In fact, a knowledge of the energy losses at the shore would be in itself a worth-while byproduct of a successful attack on the problem.

The third major difficulty which requires numerical analysis is that due to nonlinear differential equations. In fact, part of the impetus given to the construction of calculating machines during the war was due to the hope that they could be used in the solution of nonlinear differential equations. One such problem whose solution would be of immense practical importance is that of flood waves in rivers.

Suppose a river with an arbitrary channel is represented as in figure 3.3. This might be a cross section of the channel at some distance x downstream. Let $n(x,t)$ be the elevation of the water surface at some time t, and $U(x,t)$ be the horizontal velocity. Then these functions satisfy the equation

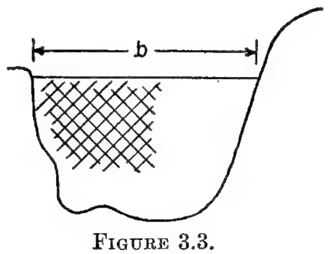

FIGURE 3.3.

$$\left.\begin{aligned} bn_t+(aU)_x&=0 \\ U_t+UU_x+gn_x&=\frac{-ca^2}{x^\alpha} \end{aligned}\right\}, \qquad (3.12)$$

where b is the surface breadth and a is the cross section of the area. The term ca^2/x^α is added in order to take into account the resistance of the channel. These equations have been known since the time of Boussinesq in the 1880's. Only in recent years has the theory been used in practice and then mostly by French hydraulic engineers. The main deterrent to the use of the theory is the enormous numerical calculation necessary to solve any but the simplest problems. It is interesting to note that the differential equations are of exactly the same type as those for the nonsteady flow of compressible gas. Actually there is a perfect mathematical analogy between the two cases in which the depth of the water in the channel plays the same role as the density of the gas. J. J. Stoker has presented a complete exposition of this analogy [5].

However, in the case of flood waves, the problems are more complicated than in gas dynamics because of the resistance of the sides of the channel and because the cross-sectional area of the water in the river changes with the depth of the water. Nevertheless, if the resistance coefficients and the cross-sectional shapes along a given river valley are known, it is possible to calculate numerically the progress of a flood wave downstream if we know the rate at which water flows into it at any point upstream. This latter figure, of course, depends upon the rainfall and many other things.

Because of the mathematical connections between these equations and those of gas dynamics, it would seem reasonable that the machines which could do one problem could do the other. A start might be made by determining the flood waves in a number of the larger and more important rivers of the United States—the Columbia, the Missouri, and the Mississippi rivers. Probably it would be found that in most of these cases, the enormous amount of data available for these rivers will still be insufficient for the purposes of a mathematical study such as is proposed. One of the fruitful effects of such a study would be to reveal the kind of data needed in order to make predictions. This in turn would help to separate the over-all problem into two parts: (1) the propagation problem itself and (2) the problem of predicting the river stage at some point upstream on the basis of the rainfall in the previous season. Another type of problem that could be studied is this: Supposing there were some dams on the river, how could we make best use of these dams in order to regulate the flow and flood waves in such rivers? Such studies could be used both in controlling flood waves and in deciding where dams should be placed.

The last type of problem to be considered is that in which the boundary conditions must be met on a free surface or on an unknown surface, such as all problems dealing with a free water surface. For example, suppose that an elevation is produced in an infinite body of water. How does this elevation progress? There is some reason to believe that a progressive wave, even in mid-ocean, will break if it goes far enough. It would be desirable to carry out the mathematical calculations and see if classical hydrodynamics agrees with this prediction.

The problem I want to consider, however, is a different one and in some respects a simpler one. It is to find the impact force on a sphere that is entering water vertically. This force is a vital consideration in the construction of seaplanes and torpedoes.

Suppose a sphere of radius A and vertical velocity B enters the water. Photographs of the entries show that the water surface behaves as in figure 3.4. Now the mathematical problem is to find a potential function $\phi(x,y,z,t)$ such that $\nabla^2\phi=0$ everywhere in the water. The boundary conditions are

$$\frac{\partial\phi}{\partial n}=-B\cos\theta$$

on the surface of the sphere where θ is the angle from the center of the sphere to any point on the sphere. Suppose that $\eta=\eta(z,t)$ is the equation of the water surface. Then there is a condition of constant pressure on the water surface, which requires that

$$\eta=\frac{1}{g}\frac{\partial\phi}{\partial t}$$

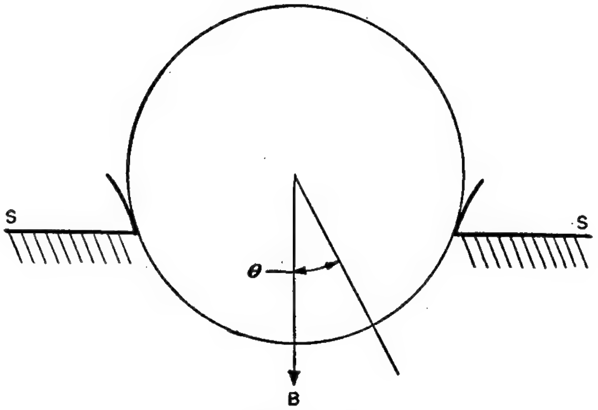

FIGURE 3.4.

on the water surface. We also have the kinematic condition

$$\frac{\partial\eta}{\partial t}=-\frac{\partial\phi}{\partial z}$$

again on the surface. Now a complete knowledge of ϕ is not needed here. We are interested only in the value of the following integral (which I shall call the induced mass given to the water by the motion of the sphere):

$$M=\rho\int\int_{W+S}\phi dx\, dy,$$

where ρ is the water density, W is the wetted portion of the sphere, and S is the water surface.

The difficulty of this problem lies in the fact that the surfaces S and W are unknown. They must be determined from the conditions of the problem. Of course, it is true that at a reasonable distance from the sphere the water surface remains practically undisturbed, but it is in the region of the sphere that the trouble occurs. From the diagram, it can be seen that there the water surface no longer has one value. As in all such problems there is a discontinuity at the highest point on the sphere which the water reaches. Even a knowledge of the type of discontinuity would be valuable.

To summarize, four typical problems that require numerical analysis have been presented. The first and last need a considerable amount of preliminary work before they are ready for solution on the machines. The second and third problems, even though they are of a spectacular type, are still of considerable scientific interest and fortunately seem to be well adapted to our present-day machines.

[1] G. N. Watson, The transmission of electric waves around the earth, Proc. Roy. Soc. A **95**, 546 (1919).

[2] M. H. L. Pryce, unpublished work.

[3] W. H. Furry, Theory of characteristic functions in problems of anomalous propagation, Radiation Laboratory Report 680, Massachusetts Institute of Technology, Cambridge, Mass. (1945).

[4] H. Lamb, Hydrodynamics, p. 334 (Cambridge University Press, Cambridge, England, 6th ed. 1932).

[5] J. J. Stoker, Formation of breakers and bores, Communications on Applied Math. **1**, 1 to 80 (1948).

4. Linear Programing

George B. Dantzig*

A certain wide class of practical problems appears to be just beyond the range of modern computing machinery. These problems occur in everyday life; they run the gamut from some very simple situations that confront an individual to those connected with the national economy as a whole. Typically, these problems involve a complex of different activities in which one wishes to know which activities to emphasize in order to carry out desired objectives under known limitations.

Consider, for example, the nutrition problem. This example apparently represents the simplest nondynamic program. Let us assume that one reason the housewife goes to the food store is to see to it that her family gets certain nutritive elements such as calories, calcium, etc. When she buys food, she does not buy a package of calcium, of calories, of vitamins, etc. What she does buy is a variety of foods, each of which contains some proportion of these elements. It is in this way that she tries to meet the daily nutritive requirements of her family [1].** There are several other considerations that also guide her selection. Thus she likes to make her choice in conformity with certain conventions and certain budgetary limitations. If the latter is important, then she may well try to minimize the cost. Mathematically we have a problem of minimizing a linear form, subject to linear inequalities (which includes equalities as a special case). The form that has to be minimized in this example is the expression totaling the price times the unknown quantity of each food purchased. Because the diet must contain so much in the way of calories, calcium, etc., one equation arises for each such nutirtive requirement. (If the requirement may be exceeded, then an inequality may arise.) One other important fact should be noted regarding the mathematical formulation of the problem: the amounts of various types of food purchased cannot be negative. It is this latter condition that makes the problem very interesting.

I wish to say a few words now about dynamic programing in the Air Forces. Let us consider the situation in which the Air Force expanded very rapidly, as in the past war. The manifold of activities that goes on in the Air Force must share in the use of a great number of different kinds of things, which I will call "equipment items." This term includes both supply and personnel. In economics the term used is "commodities." There are well over a million different kinds of supplies alone in an Air Force program. In other words, the levels of various activities, such as training, maintenance, supply, and combat must be adjusted in such a manner as not to exceed the availability of various equipment items. Indeed, activities should be so carefully phased that the necessary amounts of these various equipment items are available when they are supposed to be available so that the activities can take place.

Now, it is a legitimate question at this point to ask, "How is this to be done?" Let us consider how programing was done during the war. It took the staff well over seven months of very careful planning to come up with a program. The program thus developed was then used as a basis for action. Because of the time involved, it is clear that a careful balance was not reached on the million different items of equipment and different kinds of personnel previously mentioned. Overplanning in certain areas was necessary, and what happened, of course, was that a lot of the equipment was not used and was stored. Indeed, in any large organization there takes place a large amount of this "storage" activity that nobody cares to talk about.

Project SCOOP (Scientific Computation of Optimum Programs) is the official title of linear programing work in the U. S. Air Force. Its objective is to reduce the time it takes to plan programs. Instead of many months, we should like the planning to take a few days. Not only should we like to do the planning more quickly, but also better. This raises a question: Does there exist a general formalization of the programing problem? Perhaps human organization is such a complex web that it cannot possibly be reduced to mathematical form. It may be that programing is nothing more than a set of arbitrary decisions (I believe the usual term is "mature judgments").

On the other hand, if a systematic unified approach can be found, there is a possibility that electronic computers may be of help in speeding up programing work. For example, a representative of one of the large automobile manufacturers inquired whether the techniques being developed in the Comptroller's Office might help the automobile manufacturers in their production schedul-

* U. S. Air Force Comptroller, Washington, D. C.

**Figures in brackets indicate the literature references at the end of this paper.

ing work. The question was turned around by asking him to determine for himself whether the scheduling procedures must involve judgment at every step. For if this is true, then one can imagine how efficient it would be to start an electronic computer in operation; after about one-tenth of a second, it would have to stop, wait for some person in authority to make decisions before it could operate again for one-tenth of a second. Thus, in general, it is evident that there is no hope of using electronic computers to do planning unless one can somehow get around the large number of judgments that are required in present techniques.

Our research to date indicates that while programing work is concerned with a great variety of subjects, the underlying techniques of computations are essentially the same. It is this fact which permits the development of an orderly theory of programing that will now be discussed. I am going to present an extremely simple structure. The structure depends on a set of basic assumptions which, I am sure, will appear to you (as they did once to me) as insufficient to provide an answer. In fact, my own first reaction was one of bewilderment because there appears such an infinity of choices of levels of activities that can take place in a large organization.

If one views the general structure as a complex of activities which share in the use of items of equipment, then activities can be grouped in many ways consistent with the availability of equipment. It is clear that the particular grouping over time of activities to form a program depends on the objectives of the organization or more generally the "economy" under discussion. Accordingly, the basic problem of programing is to construct a program of activities which is consistent within itself and which maximizes, in some sense, the objectives of the economy. We shall approach the problem of constructing a mathematical model by assuming that the amounts and kinds of equipment required to carry on the activities are known and that the objectives can be stated in quantitative terms.

Notation.—Let $0 \leq t \leq T$ be the time span of the program. Each activity, if it takes place over several time periods $0 \leq t \leq 1$, $1 \leq t \leq 2$, . . . , $T-1 \leq t \leq T$, is broken down into a set of subactivities, each assigned to one of these time periods. Let us denote the j^{th} type of activity in a typical time period by A_j, where $j=1, 2, \ldots, n$. (A superscript t will denote where necessary the t^{th} period, $t=1, 2, \ldots, T$). Let the quantity of the i^{th} type of equipment item required to be on hand at the beginning of the period to carry on the j^{th} activity be noted by E_{ij} where $i=1, 2, \ldots, m$. At the end of the period of time the amount of this equipment item will almost always undergo change. In a sense, the activity A_j can be thought of as operating on E_{ij}, transforming it to a quantity $\overline{E}_{ij}$ by the end of the period. In other words, E_{ij} becomes $\overline{E}_{ij}$ when operated on by the j^{th} activity. For example, in the training activity in the Air Force one starts out with advanced student pilots; at the end of a period of say a month or 6 weeks, they will become full-fledged pilots; so that, if $i=1$ represents students and $i=2$ pilots, then $E_{1j} \rightarrow \overline{E}_{1j}=0$ and $E_{2j}=0 \rightarrow \overline{E}_{2j}$.

Basic Assumptions.—Next let us examine certain assumptions underlying a linear structure. First, *the assumption of additivity* of certain subsets relative to an equipment item. This statement on additivity is simply a bookkeeping assumption. It states that if one takes the quantity of equipment and assigns it to two activities, the resultant quantity of equipment will be the sum of the quantity of equipment assigned to each of the activities. There is no overlapping: common equipment or common facilities are not used. On paper, at least, the equipment is broken up and part is assigned to each activity. The second assumption concerns *completeness* of activities: it states that the totality of equipment on hand at time $t-1$ is equal to the sum assigned to different activities during the t^{th} period,[1] i. e.,

$$E_i^{(t-1)} = \sum_{j=1}^{n} E_{ij}^{(t)}, \qquad (i=1, 2, \cdots, m).$$

Thus if one has a complete list of all of the activities and totals the amounts of an equipment item assigned to each one, the sum equals the total amount of the equipment. The third assumption is also concerned with the completeness of activities; it states that the total amount of equipment on hand at the end of the t^{th} period (time t) is equal to the total amount produced by the various activities during the t^{th} period, i. e.,

$$E_i^{(t)} = \sum_{j=1}^{n} \overline{E}_{ij}^{(t)}, \qquad (i=1, 2, \cdots, m).$$

Now let us consider the more important assumptions on proportionality. Assumption four: the amount of equipment required to carry on the activity is proportional to the level of the activity. To illustrate, if the activity is building an electronic computer, then the assumption states that the building of two electronic computers instead of one electronic computer will require twice as many tubes, twice as many mercury delay lines, etc. In theory, at least, if one has twice as many of these components at the beginning of the period, twice as many electronic computers will also be produced, This leads to assumption five: the amount of equipment produced by an activity is proportional to the level of the activity. The equations expressing proportionality are given by

$$E_{ij}^{(t-1)} = \alpha_{ij} \cdot X_j^{(t)}$$

$$\overline{E}_{ij}^{(t)} = \overline{\alpha}_{ij} \cdot X_j^{(t)}$$

[1] The t^{th} time period extends from $t-1$ to t. The superscript in $E_{ij}^{(t)}$ and $\overline{E}_{ij}^{(t)}$ refers to equipment assigned to or produced by the j^{th} activity in the t^{th} time period, whereas in $E_i^{(t)}$ it refers to total equipment on hand at time t.

where $X_j^{(t)} \geq 0$ represents the level or quantity of activity during the t^{th} period and α_{ij} and $\overline{\alpha}_{ij}$ (called input and output coefficients, respectively) are the coefficients of proportionality referred to in assumptions four and five.

Based on these five assumptions, the equations of the dynamic system may easily be derived. If $E_i^{(0)}$ represents given initial conditions, then the levels of activities during the first time period must satisfy the set of equations:

$$E_i^{(0)} = \sum_{j=1}^{n} \alpha_{ij} \cdot X_j^{(1)}, \qquad (i=1, 2, \cdots, m)$$

The levels of activity during the first time period determine the amount of each equipment item available for operations during the second time period; thus

$$E_i^{(1)} = \sum_{j=1}^{n} \overline{\alpha}_{ij} \cdot X_j^{(1)} = \sum_{j=1}^{n} \alpha_{ij} \cdot X_j^{(2)}.$$

Similarly, at the beginning of the $(t+1)$ period,

$$E_i^{(t)} = \sum_{j=1}^{n} \overline{\alpha}_{ij} \cdot X_j^{(t)} = \sum_{j=1}^{n} \alpha_{ij} \cdot X_j^{(t+1)},$$

where $i=1, 2, \ldots, m$.

Matrix Notation.—Let $X^{(1)}$ represent the vector of activities that take place in the first time period, and let $\alpha = [\alpha_{ij}]$ be the matrix of input coefficients and $\overline{\alpha} = [\overline{\alpha}_{ij}]$ be the matrix of output coefficients, then

$$\begin{aligned} E^{(0)} &= \alpha X^{(1)} \\ 0 &= -\overline{\alpha} X^{(1)} + \alpha X^{(2)} \\ 0 &= \qquad\qquad -\overline{\alpha} X^{(2)} + \alpha X^{(3)} \\ &\cdot \\ &\cdot \\ &\cdot \end{aligned}$$

The terms "input" and "output" are borrowed from similar terms used in a model constructed by W. W. Leontief for describing the structure of the American Economy [2]. The model described here was developed by the author as a generalization of the Leontief model to a dynamic situation. However, in the generalized form it is more closely analogous to a dynamic set of equations developed by J. von Neumann in 1932 [3].

The Maximizing Function.—The system of equations is subject to the side condition that the levels of activities, comprising the elements of the vectors $X^{(t)}$, are nonnegative, i. e.

$$X_j^{(t)} \geq 0$$

Usually, in spite of this restriction, there are many possible solutions to the system when $X^{(t)}$ are regarded as unknown vectors. This is not surprising when one reflects that there are many possible programs (i. e. choices of $X^{(t)}$ consistent with initial status. Naturally, not all of these programs make much sense in terms of the "objectives" of the organization or composite activity. Thus in the case of the housewife, after imposition of additional restrictions that reflect conventions and preferences, there may be some degrees of freedom left. In this case her objective will be to minimize the cost of the diet. On the other hand, the objectives of the Air Force might be to fly as many sorties as possible in the event of war. If we think of certain of the X's as representing combat activities given in sortie units, the objective might be to maximize a certain linear function of the $X_j^{(t)}$ which evaluates the total sorties flown during the time span of the program. As a third example, in a large business organization making automobiles, many of the X's consist in the production of various types of automobiles, buying parts, etc. There are many alternative actions that may be taken, but that which yields maximum profits to the enterprise is apt to be the most significant one. All activities of the enterprise will have costs associated with them. Those concerned, however, with the selling of cars will have a positive output of money. Again, the objective function can be expressed as linear function of various activities. The best program in this case is the one which maximizes profits.

Let me say a few words about the techniques by which we plan to maximize a linear form subject to linear restrictions. We shall discuss the techniques of maximization in terms of a more general mathematical problem: maximize a linear form of nonnegative variables $X_1, X_2, \ldots, X_n$, i. e.,

$$X_1.b_1 + X_2.b_2 + \ \ldots\ + X_n.b_n = \max, \qquad (X_j \geq 0)$$

where X_j satisfy a system of m equations, in vector form

$$X_1 \cdot P_1 + X_2 \cdot P_2 + \ \ldots\ + X_n \cdot P_n = P_0,$$

where P_j has coordinates $(a_{1j}, a_{2j}, \ldots, a_{mj})$.

In order to arrive at an intuitive geometrical picture, let us suppose that it is known that the sum of the X's is equal to unity. We could then think of the X's as nonnegative weights assigned to a set of points in say m dimensional space [2] and form a linear combination of these points to produce a given point P_0. In other words, the point P_0 is a center of gravity of known points $P_1, P_2, \ldots, P_m$ with unknown weights $X_1, X_2, \ldots, X_n$. If we now add a b-coordinate to each of the points, $P_1, P_2, \ldots$ where $b_1, b_2, \ldots$ is taken from the maximizing form, then our problem is to create a center of gravity lying on a

[2] Strictly speaking, the condition $\Sigma X_i = 1$ substitutes for one of the equations hence one of the coordinates can be dropped and the point *Pi* plotted in $m-1$ dimensional space.

given line, P_0, parallel to the b-axis whose b-coordinate is maximum.

To illustrate, let us examine figure 4.1: If we think of these points as spanned by a convex, then we are looking for a point where this line pierces an "upper" face of the convex. In the diagram (P_4, P_5, P_6) represents the top face.

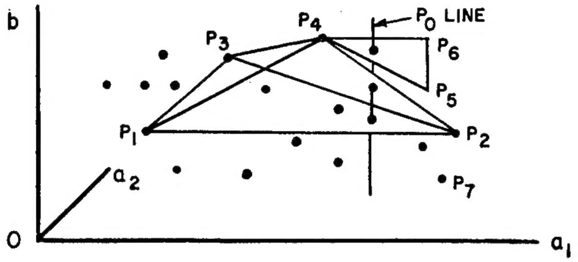

FIGURE 4.1.

It can be shown that the solution will embody the following characteristics: That all points (activities) will be given a weight of zero except m out of n of them, where m equals the number of equations. The remaining m will be given nonnegative weights corresponding directly to the level of activity. One method of solving the problem might be as follows: Assume that we have some kind of a program that means we have found some combination of points that generates P_0 with nonnegative weights, e. g., P_1, P_2, P_3 in the diagram. Now we wish to improve the "program", i. e., find a triangle that cuts P_0 in a higher point. To do this, we consider any point P_i that lies above the plane of the triangle P_1, P_2, P_3, for example, P_4. We join such a point to P_1, P_2, P_3 forming a simplex in space. The line P_0 pierces the simplex in two points. One point represents the b of the solution involving P_1, P_2, P_3, the other point is higher, since all points in the simplex lie above the face P_1, P_2, P_3. Thus a solution involving just P_2, P_3, P_4 represents an improvement over P_1, P_2, P_3. It is clear that by iterating this procedure one will eventually obtain the best program in a finite number of iterations. This method of solution is known as the "simplex" technique.

There have been several other techniques suggested by J. von Neumann and others. There is a very close relationship between this problem and the problem of determining the optimum mixed strategy in game theory. In fact, the problem here presented includes the game problem as a special case. It is more difficult to show that the game problem is actually equivalent to the program problem.

To date procedures that have been devised for maximizing a linear form subject to linear restrictions involve, even for the simplest problems, a great number of computations. With the advent of high speed electronic computers many everyday programing problems which involve alternative choices of action may be expected to be solved by techniques outlined here. Very large scale programing problems, however, may have to wait for electronic computers considerably faster than those currently contemplated. Research in the field may be said to hardly have started. Faster computational techniques are needed and these can only be arrived at by a better understanding of the problem. The linear structure itself, of course, is only a starting point. Computational methods will have to be extended to nonlinear areas also.

[1] G. J. Stigler, The cost of subsistence, J. Farm. Econ. **27**, 303 to 314 (1945).

[2] W. W. Leontief, Structure of the American economy 1929–1929 (Harvard University Press, Cambridge, Mass., 1941).

[3] J. von Neumann, A model of a general economic equilibrium, Review of Econ. Studies, XIII (1), 1 to 9 (1945–46).

[List continued from cover page 2]

MT17. Miscellaneous Physical Tables: Planck's Radiation Functions and Electronic Functions (Originally published in J. Opt. Soc. Am., Feb. 1940.) (1941) vi+58p. $1.50.

MT18. Table of the Zeros of the Legendre Polynomials of Order 1–16 and the Weight Coefficients for Gauss' Mechanical Quadrature Formula. (Reprinted from Bul. Am. Math. Soc., Oct. 1942.) 5p. 25¢.

MT19. On the Function $H(m,a,x)=\exp(-ix)\times F(m+1-ia, 2m+2; 2ix)$: Includes table of $H(m, a, x)$ and its first derivative. (Originally published in J. Math. Phys., Dec. 1942; reissued by NBS 1949.) 20 p. 10¢.

MT22. Table of Coefficients in Numerical Integration Formulae. 1949. 2p. 5¢.

MT24. Coefficients for Numerical Differentiation With Central Differences. (Reprinted from J. Math. Phys., Sept. 1943.) 21p. 25¢.

MT25. Seven-Point Lagrangian Integration Formulas. (Originally published in J. Math. Phys., Dec. 1943; reissued by NBS 1949.) 4p. 10¢.

MT26. A Short Table of the First Five Zeros of the Transcendental Equation $J_0(x)\,Y_0(kx)-J_0(kx)\,Y_0(x)=0$. (Reprinted from J. Math. Phys., Dec. 1943.) 2p. 25¢.

MT27. Table of Coefficients for Inverse Interpolation With Central Differences. (Originally published in J. Math. Phys., Dec. 1943; reissued by NBS 1949.) 15p. 25¢.

MT29. Table of Coefficients for Inverse Interpolation With Advancing Differences. (Reprinted from J. Math. Phys., May 1944.) 28p. 25¢.

MT30. A New Formula for Inverse Interpolation. 1949. 4p. 5¢.

MT31. Coefficients for Interpolation Within a Square Grid in the Complex Plane. (Reprinted from J. Math. Phys., Aug. 1944.) 11p. 25¢.

MT32. Table of Coefficients for Differences in Terms of the Derivatives. (Reprinted from J. Math. Phys., Nov. 1944.) 4p. 25¢.

MT34. Inverse Interpolation for Eight-, Nine-, Ten-, and Eleven-Point Direct Interpolation. (Reprinted from J. Math. Phys., May 1945.) 4p. 25¢.

MT35. Table of Coefficients for Double Quadrature Without Differences, for Integrating Second-Order Differential Equations. (Reprinted from J. Math. Phys., Nov. 1945.) 6p. 25¢.

MT36. Formulas for Direct and Inverse Interpolation of a Complex Function Tabulated Along Equidistant Circular Arcs. (Reprinted from J. Math. Phys., Nov. 1945.) 3p. 25¢.

MT37. On the Computation of Mathieu Functions. 1949. 20p. 25¢.